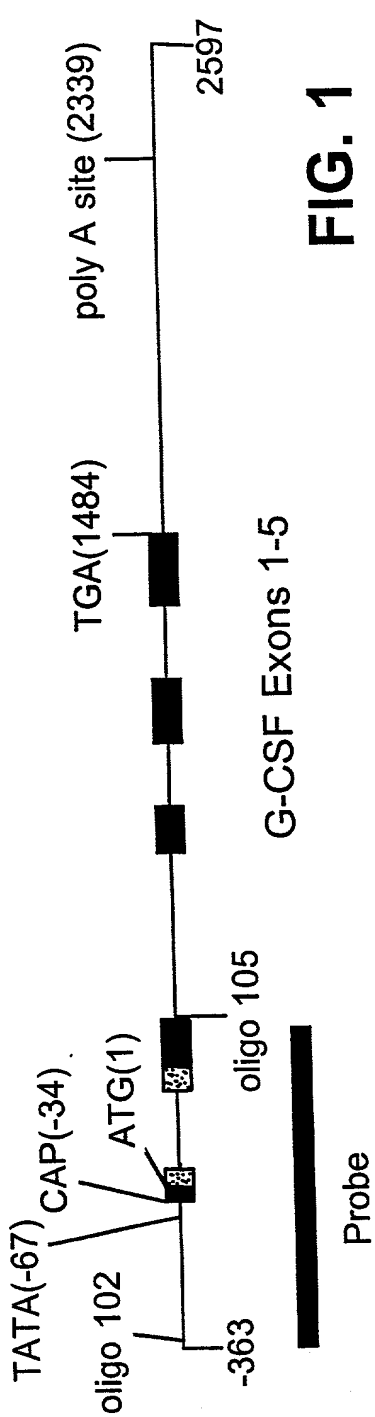
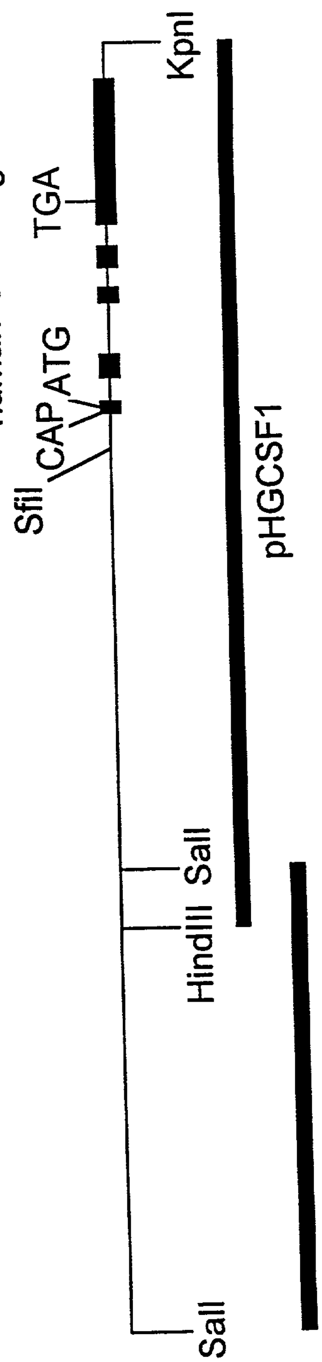


poly A site (2339)
2597
TGA(1484)
G-CSF Exons 1-5
TATA(-67)
CAP(-34)
ATG(1)
oligo 102
-363
oligo 105
Probe

FIG. 1



human G-CSF gene
KpnI
TGA
ATG
CAP
SfiI
pHGCSF1
SalI
HindIII
SalI
pHGCSF4

FIG. 2

SalI (-6596)

-6597 GTCGACCTGC AGGTCAACGG ATCACTTGAG GACAGTAGTT CAAGACCAGC CTGGGCAGCA TAGGGAGACT GTCTCTACGA AAAATCAAAA AATTATGGCC
-6497 GGGCATGGTG GCTCACGTCT GTAATCCCTG AACTTTGGGA CATCAAGGCA AGTGGATCAC TTGAGGTCAG GAGTTCGAGA CTAGCCTGGC CAACATGGTG
-6397 AAACCCTATC TCCACTAAAA AATACAAAAA TTAGCCAGGC ATGGTGGCAG GCACCTGTAA TCCCGGCTAC TCAGGAGGCT GAGGCAGGAG AATCACTTGA
-6297 ACCCAGGAGG CGGAGGTTGC AGTGAGCTGA GATCACACCA CTGCACTCCA GCCTGGGTGA CAGAGCAAGA CTCTATCTCA AAAAAAATAA AAAAATAAAA
-6197 AAATTAGCCA GGCATGGTAG TGCACACCTC TAGTCTCAGC TACTCAGGAG GCTGAGGTGG GAGGATCACT TGAACCTGGG GCAGTCAAGG CTACAGTGAG
-6097 CCAAGATCAT GCCACTACAC TCCAGCCTGG GCAACAGAGA GAGACCCTGT CTCTAAAAAA ATAATAATAA TAAAGAAAAA AACAGCTCTG TTTATGTCTC
-5997 CTGGTCCATA CATACTACTA TGTATATAGT TTGCAAACTC AAAGATCCAG ATAGTCAATT TTTTAGGCTT GTGGGCCGTA TGGTCTCTGT CACAATCACT
-5897 CTGCCCTGTC TTTCTAGCAC AAAAGCAGCT ATAAACAATA CATACATGAA TTTTTTATAG ACATCGAGAT TTGAATTTCA TATGATTTTT ACATTTTATA
-5797 AAATAATCTT TTTAAAAATT TTCCCCTAAC CATTTAAAAG TGTAAAAGCC GGCCAGCGCG CCATCGTCAC GCCTGTAATT CCAGCACTTT GGGAGGCTGA
-5697 GGTGGGCAGA TCACTTGAGA TCAACAGTTC GAGACCAGCC TGGCCAACAT AGCAAAACCC CATTTCTACT AAAAATAAAA AAATTAGCTG GGCATAGTGG
-5597 TGCACACCTG TGATCCCAGC TACTTGGGAG GCTGAGGCAG GAGAATCGCT TGAACCTGGG AAGCGGAGGT TGCAGTGAGC CAACATCATG CCACTGCACT
-5497 CCAGCCTGGG TGACAGAGTG AGACTTCGTC TCAACGAAAA AAAAAAGTGT AAAAGCCATT CCTAATTCAG TGTACATCAG TGTACATACT CAGGTCTGCG
-5397 TACTCCTGCT CTGAGGCATA CCTGAGAAGT AGAGTTGCTT GGTCACAGGA CATACACATT TCCACATTAA CTAGACACTA CCAAGTTGCC ATCCAAGGAG
-5297 GTTTTTTTTT TACAATCTAC ACTCCCCCCA GCAACAAATG AGAGTTACTC CAGATCCTTT ACAAAGATGC TCTAAGCCCA GTACCAGATG AAAACAGGAA
-5197 GTGGGAGGGG AAGCTGCCAG CCCCTTCTAA CCATGAAGAA ATACCTGGTA GAGCCTTCTG GATGCTGGAA GGATGAATAA CGGGGGTCTC TGGAGCCTGC
-5097 CCCCTGTCAG ATCACTGTGA CTTCTGAGCC TCCAGTCCAG TCTCAGCCCC ATGTGTCATG GCCAGTGATA ATGAGCCCTC ACTCTCTGTT TGGTCTTTAT
-4997 TCTCCCCATG TGGGGCTGAA GTCTGGATTG AGCCGTTATT CAAGATGTAC AGCTTTCTTG ACAGGAAAGT AGTGTCACAG AAACAGCAGG GGCTTGGCAA
-4897 GATGATCTAA CTGCAAATCC TACCTGGCTC AGCCACCAGC TAGTTCTGTG ATCTTGAACA AGTTTTTTCA CTTCTCTGAG GCCATCCCTT GGCTACAACA
-4797 CACCAGTTGG TTGACAGGAT GAAATGACGA AGTCCCTTAC ACCTGTAATC CCAGCACTTT GGGAGGCCAA GGCGGGTGGA TGGCTTGAGC CTGAGAGGTG

SphI (-4693)

-4697 ACAGCATGCC GGCAGTCCTC ACAGCCCTCG TTCGCTCTCG GCGCCTCCTC TGCCTGGGCT CCCACTTCGG TGGCACTTGA GGAGCCCTTC AGCCCACCGC
-4597 TGCACTGTGG GAGCCCCTTT CTGGGCTGGC CAAGGCCAGA GCCGGCTCCC TCAGCTTGCA GGGAGGTGTG GAGGCAGAGG CTCAAGCAGG AACCGGGGCT

SmaI (-4406)

-4497 GCGCACGGCG CTTGCGGGCC AGCTGGAGTT CCGGGTGGGC GTGGGCTTGG CGGGCCCCGC ACTCGGAGCA GCGGGCCAGC CCTGCCAGGC CCCGGGCAAT

FIG. 3A

SmaI (-4383)

-4397 GAGAGGCTTA GCACCCGGGC CAGCGGCTGC GGAGGGTGTA CTGGGTGCCC CAGCAGTGCC AGCCCGCCGG CGCTGTGCTC GCTCGATTTC TCACTGGGCC
-4297 TTAGCAGCCT TCCCGCGGGG CAGGGCTCGG GACCTGCAGC CCGCCATGCC TGAGCCTCCC CTCCATGGGC TCCTGTGCGG CCCGAGCCTC CCCGACGAGC
-4197 ACCACCCCCT GCTCCACAGC GCCCAGTCCC ATCGACCACG CAAGGGCTGA GAAGTGCGGG CGCACGGCAC CGGGACTGGC AGGCAGCTAC CCCTGCAGCC
-4097 CTGGTGCGGA ATCCACTGGG TGAAGCCAGC TGGGCTCCTG AGTCTGGTGG AGACTTGGAG AACCTTTATG TCTAGCTCAG GGATCGTAAA TACACCAATC
-3997 AGCACCCTGT GTCTAGCTCA GGGTCTGTGA ATGCACCAAT CCACACTCTG TATCTAGCTA CTCTGATGGG GCCTTGGAGA ACCTTTATGT CTAGCTCAGG
-3897 GATTGTAAAT ACACCAATCG GCACTCTGTA TCTAGCTCAA GGTTTGTAAA CACACCAATC AGCACCCTGT GTCTAGCTCA GGGTATGTGA ATGCACCAAT

HindIII (-3722)

-3797 CGACAGTCTG TATCTGGCTA CTTTCATGGG CATCCGTGTG AAGAGACCAC CAAACAGGCT TTGTGTGAGC AATAAAGCTT CTATCACCTG GGTGCAGGTG
-3697 GGCTGAGTCC GAAAAGAGAG TCAGCGAAGG AGATAAGGGT GGGGCCGTTT TATAGGATTT GGGTAGGTAA AGGAAAATTA CAGTCAAAGG GGGTTTGTTC
-3597 TCTGGCCGGG CAGGAGTGGG GGGTCGCAAG GTGCTCAGTG GGGGTGCTTT TTGAGCCAGG ATGAGCCAGG AAAAGGACTT TCACAAGGTA ATGTCATCAA
-3497 TTAAGGCAAG GACCCGCCAT TTACACCTCT TTTGTGGTGG AATGTCATCA GTTAAGTTTC GGGGCAGGGC ATATTCACTT CTTTTGTGAT TCTTCAGTTA
-3397 CTTCAGGCCA TCTGGGCGTA TATGTGCAAG TTACAGGGGA TGCGATGGCT TGGCTTGGGC TCAGAGGCTT GACAGCTACT CTGGTGGGGC CTTGGAGAAT

SalI (-3290)

-3297 GTTTGTGTCG ACACTCTGTA TCTAGTTAAT CTAGTGGGGA CGTGGAGAAC CTTTGTGTCT AGCTCAGGGA TTGTAAACGC ACCAATCAGC GCCCTGTCAA
-3197 AACAGACCAC TCGGCTCTAC CAATCAGCAG GATGTGGGTG GGGCCAGATA AGAGAATAAA AGCAGGCTGC CCGAGCCAGC AGTGGCAACC GCACAGGTCC
-3097 CTATCCACAA TATGGCAGCT TTGTTCTTTT GCTGTTTGCG ATAAATCTTG CTACTGCTCG CTTTTTGGGT CCACACTGCT TTTATGAGCT GTAACACTCA
-2997 CCACGAAGGT CTGCAGCTTC ACTCCTGAAG CCACTAAGAC CACGAGCCCA CCGGAGGAAT GAACAACTCC GGCCGCGCTG CCTTAAGAGC TATAACACTC
-2897 ACCGCGAAGG TCTGCAGCTT CACTCCTCAG CCAGCGAGAC CACGAACCCA CCAGAAGGAA GAAACTGCGA ACACATCTGA ACATCAGAAG GAACAAACTC
-2797 CAGATGCACC ACCTTAAGAG CTGTAACACT CACTGCGAGG GTCCGCGGCT TCCTTCTTGA AGTCAGTGAG ACCAAGCACT CACCAGTTTC GGACACAAGC
-2697 CCAGGAGTTT GAGATCAGCC TGGGCAACAT GATGAAATGC CCTCTCTGCA AAAAAAAAAA AAATTACAAA AATTGGCGGA GTCGGTGGTC CGTGGCCTGT
-2597 GGTCCCAGCT ACGCGGAGGC TAAAGTTGGG AGGATCGCTT GAGCCTGGGA GGTGAAGACT GCAGTGAGCT GTGATTGTAC CACAGCCCTC TAGGCTGGGG
-2497 GACAGACTGA GACCCTGTTT CCCCTCCGCA AAAAAATTGA CAAAAGTGTA ATAAGAGGTG CCTGATATGG CTAGGTGCAG TGGCTCATGC CTGTAATCCC
-2397 AGCACTTTGG GAAGCCGAGG CGGGCGGGTC ACCTAAGGTC AGGAGTGTGA GACCAGCCTG GCCAACATGG AGAAAGCCCA TCTCTTCTAA AAATACAAAA

SphI (-2269)

-2297 TTAGCCGGCT GTGGGGGCAG TGGTGGAGCA TGCCTGTAAT CCCAGCTACT CAGGAGGCTG AGGCAGGAGA ATCACTTGAA CCCAGGAGGC GGCGGTTGCA
-2197 GTGAGCCGAG ATCGTGCCAT TGCACTCCAC CCACTCCAGC CTGGGCAACA AGAGCCAAAC TCTGTCTTAA AAAAAAAAAA AAAGTGCCTG ACATATAAGA
-2097 GGTGTGCAAT GCATAGTTGC CAGGCAACAT GTTTAAGAAT GTGGAGCTCC TGCCTTCCAT GGTCCTGTTA AAAACCCACC CTCAAGGCCA GGTGCAGTGG
-1997 CTCATGCCTA TAATCCCAGC ACTTTGGGAG GCCGAGGCGG GTGGATCACC TGAGGTCAGG AGTTCGAGAC CAGCCTGACC ACCAACATGG TGAAATCCCA

FIG. 3B

```
                                  SphI (-1858)
-1897 CCTCTACTAA AAATACAAAA TTAGATGAGC ATGGTGGTGC ATGCCTGTAA TCCCACCTAC TGGGAGGCTG AGGCAGGAAA ATCACTAGAA CCAGGGAGGC
-1797 GGAGGTTGTA GTGAGCCGAG ATCGTGCCAT TGCACTCCAG CCTGAGCAAT GAGCGAAACT CCATCTCAAA AAAACAACAA CAAAAACCCA CTCTCTACTC
-1697 CAGGGAGCTG GGTACAGAGC TGGGCCACAT CAGTGCAAGG TGCTGAGCCA CAGAGCTAAG GCGACGTGCA GGACCGCGGA CCAGATAACA GTGTGTGAGA
-1597 TCAGTGTGTG AGATCAGACG TCCCTGCCAT TGGTGACCAC CAGGGGGCCC CCAAGCACCA GAGATGGCCC CATCCAGTCA CCACATCCAC TTCTCATCCA
-1497 GAGATGTCTG TTTCTTGGCA CGCTGGGGTA AATTAGGACA GAAGGTGACA GTCTTGGGTG TGGTCAGTCA GACTGCCCCA GGCAGGCCTT GTGGCTGTAG
-1397 AAAACGTTCA GGCCTAGGCG CGACGTGGCT CACGCCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGCGGGT GGATCACGAG GTCAGGAGAT CGTGACCATC
-1297 CTGGCTAACA CGGTGAAACC CCGTCTCTAC TAAAAATACA AAAAATTGGC CGGGCATGGT GGCGGGCACC TGTAGTTCCA GCTACTCGGG AGGCTGAGGC
-1197 AGGAGAATGG GTGAACCCGA GAGGCAGAGT TTGCAGTGAG CCGAGATCGC GCCACTGCAC TCCAGCCTGG GCGACAGAGC AAGACTCCAT CTGGAAAAGA
-1097 AAAAGAAAAC GTTCAGGTCT GAGCCAGAGG CCCAGGCTGT AATTCTGTCA CTTACCATGA CCTTGGGCAA GGCACTTCCT TCCCTGGCCA GTTCACGGGG
 -997 TTGGAATCGA CTCCAAGGTC CCTTCCAGCA TTAACGCTGC ATGGTTCTAA GATGAGAAGA TGGGGCAGTT TCCCCTCTCT CACCCCAGCC CGTGTCCACT
 -897 TCAAGGTGAA TGACCAGGGA AGTCACGTGT CCCAATCCCG CAGTTCCAAA GCCCTTGGGA CCCTACTGTC AGGGTCGTGC ACGAGGAGGT GAAGGTCAGG
 -797 TGAGCCAATC GCCTCGAAGG GTCTTGCCTC ATTCGGGACA GACATCCGGT TTCCTCTGGC TCTACCGGAT TCTAGGGCTT TAGCCGAATG AGTCATGGGG
 -697 GCGGGGGGTT TCTGGGGAGT TCCCAGCTAA TCAACTTGGA CAGACAGCCT GGAACTTTCG ATGGTGCCTA TCCAAGTGTG GGGTGGGCAC AGCAGCCAAG
 -597 ACCCAATGTC CTTATCTCAG GTAGGGGCTC AGGAGGTCTC CCAGACAGGC AGCCTCCGGA GAGTTTGGGG GTAGGAATGG GAGCAACCAG GCTTCTTTTT
 -497 TTCTCTCTTA GAATTTGGGG GCTTGGGGGA CAGGCTTGAG AATCCCAAAG GAGAGGGGCA AAGGACACTC CCCCACAAGT CTGCCAGAGC GAGAGAGGGA
                                                                                                          SfiI (-297)
 -397 GACCCCGACT CAGCTGCCAC TTCCCCACAG GCCTCTGCCG CTTCCAGGCG TCTATCAGCG GCTCAGCCTT TGTTCAGCTG TTCTGTTCAA ACACTCTGGG
 -297 GCCATTCAGG CCTGGGTGGG GCAGCGGGAG GAAGGGAGTT TGAGGGGGGC, AAGGCGACGT CAAAGGAGGA TCAGAGATTC CACAATTTCA CAAAACTTTC
 -197 GCAAACAGCT TTTTGTTCCA ACCCCCCTGC ATTGTCTTGG ACACCAAATT TGCATAAATC CTGGGAAGTT ATTACTAAGC CTTAGTCGTG GCCCCAGGTA
                              TATA box (-67)                         CAP (-34)                               ATG (1)
  -97 ATTTCCTCCC AGGCCTCCAT GGGGTTATGT ATAAAGGGCC CCCTAGAGCT GGGCCCCAAA ACAGCCCGGA GCCTGCAGCC CAGCCCCACC CAGACCCATG
                                                                                                          1▸Met
                                    intron 1 (41)
    4 GCTGGACCTG CCACCCAGAG CCCCATGAAG CTGATGGGTG AGTGTCTTGG CCCAGGATG (SEQ ID NO: 1)
   2▸AlaGlyProA laThrGlnSe rProMetLys LeuMet (SEQ ID NO: 2)
```

FIG. 3C

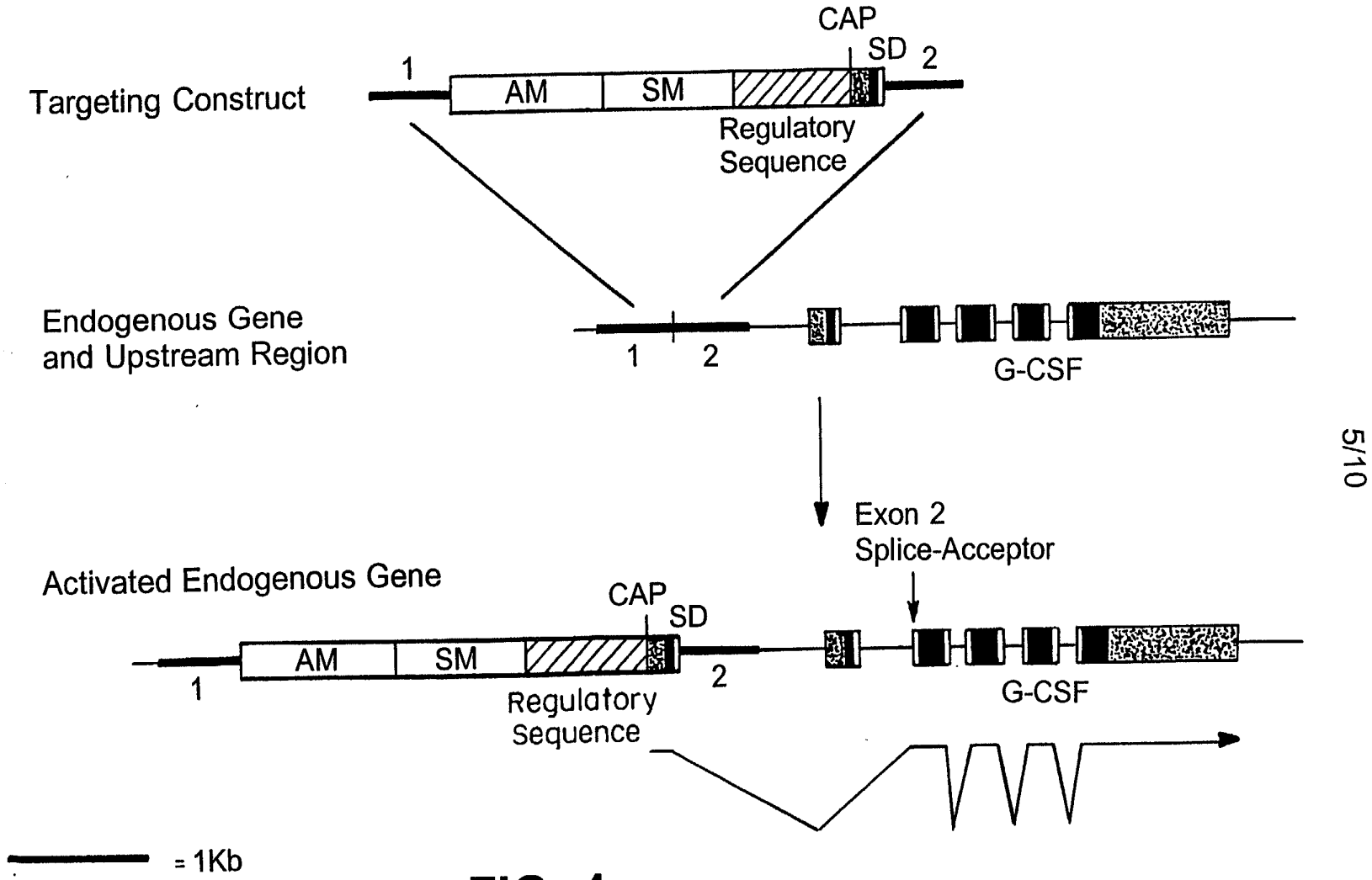

Targeting Construct
CAP
SD
1
2
AM
SM
Regulatory
Sequence
Endogenous Gene
and Upstream Region
1
2
G-CSF
Exon 2
Splice-Acceptor
Activated Endogenous Gene
CAP
SD
AM
SM
1
2
Regulatory
Sequence
G-CSF
= 1Kb

FIG. 4

GATCACTTGAGGACAGTAGTTCAAGACCAGCCTGGGCAGCATAGGGAGACTGTCTCTACGAAAAA
TCAAAAAATTATGGCCGGGCATGGTGGCTCACGTCTGTAATCCCTGAACTTTGGGACATCAAGGC
AAGTGGATCACTTGAGGTCAGGAGTTCGAGACTAGCCTGGCCAACATGGTGAAACCCTATCTCCA
CTAAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGCACCTGTAATCCCGGCTACTCAGGAGGC
TGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATCACACCACTGC
ACTCCAGCCTGGGTGACAGAGCAAGACTCTATCTCAAAAAAAATAAAAAAATAAAAAAATTAGCC
AGGCATGGTAGTGCACACCTCTAGTCTCAGCTACTCAGGAGGCTGAGGTGGAGGATCACTTGAA
CCTGGGGCAGTCAAGGCTACAGTGAGCCAAGATCATGCCACTACACTCCAGCCTGGGCAACAGAG
AGAGACCCTGTCTCTAAAAAAATAATAATAATAAAGAAAAAAACAGCTCTGTTTATGTCTCCTGG
TCCATACATACTACTATGTATATAGTTTGCAAACTCAAAGATCCAGATAGTCAATTTTTTAGGCT
TGTGGGCCGTATGGTCTCTGTCACAATCACTCTGCCCTGTCTTTCTAGCACAAAAGCAGCTATAA
ACAATACATACATGAATTTTTTATAGACATCGAGATTTGAATTTCATATGATTTTTACATTTTAT
AAAATAATCTTTTTAAAAATTTTCCCCTAACCATTTAAAAGTGTAAAAGCCGGCCAGGGCGCCAT
CGTCACGCCTGTAATTCCAGCACTTTGGGAGGCTGAGGTGGGCAGATCACTTGAGATCAACAGTT
CGAGACCAGCCTGGCCAACATAGCAAAACCCCATTTCTACTAAAAATAAAAAAATTAGCTGGGCA
TAGTGGTGCACACCTGTGATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGG
GAAGCGGAGGTTGCAGTGAGCCAACATCATGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGAC
TTCGTCTCAACGAAAAAAAAAAAGTGTAAAAGCCATTCCTAATTCAGTGTACATCAGTGTACATAC
TCAGGTCTGCGTACTCCTGCTCTGAGGCATACCTGAGAAGTAGAGTTGCTTGGTCACAGGACATA
CACATTTCCACATTAACTAGACACTACCAAGTTGCCATCCAAGGAGGTTTTTTTTTACAATCTA
CACTCCCCCCAGCAACAAATGAGAGTTACTCCAGATCCTTTACAAAGATGCTCTAAGCCCAGTAC
CAGATGAAAACAGGAAGTGGGAGGGGAAGCTGCCAGCCCCTTCTAACCATGAAGAAATACCTGGT
AGAGCCTTCTGGATGCTGGAAGGATGAATAACGGGGTCTCTGGAGCCTGCCCCCTGTCAGATCA
CTGTGACTTCTGAGCCTCCAGTCCAGTCTCAGCCCCATGTGTCATGGCCAGTGATAATGAGCCCT
CACTCTCTGTTTGGTCTTTATTCTCCCCATGTGGGGCTGAAGTCTGGATTGAGCCGTTATTCAAG
ATGTACAGCTTTCTTGACAGGAAAGTAGTGTCACAGAAACAGCAGGGGCTTGGCAAGATGATCTA
ACTGCAAATCCTACCTGGCTCAGCCACCAGCTAGTTCTGTGATCTTGAACAAGTTTTTTCACTTC
TCTGAGGCCATCCCTTGGCTACAACACACCAGTTGGTTGACAGGATGAAATGACGAAGTCCCTTA
CACCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGATGGCTTGAGCCTGAGAGGTGACAG
CATGCCGGCAGTCCTCACAGCCCTCGTTCGCTCTCGGCGCCTCCTCTGCCTGGGCTCCCACTTCG
GTGGCACTTGAGGAGCCCTTCAGCCCACCGCTGCACTGTGGGAGCCCCTTTCTGGGCTGGCCAAG
GCCAGAGCCGGCTCCCTCAGCTTGCAGGGAGGTGTGGAGGGAGAGGCTCAAGCAGGAACCGGGGC
TGCGCACGGCGCTTGCGGGCCAGCTGGAGTTCCGGGTGGGCGTGGGCTTGGCGGGCCCCGCACTC
GGAGCAGCGGGCCAGCCCTGCCAGGCCCCGGGCAATGAGAGGCTTAGCACCCGGGCCAGCGGCTG
CGGAGGGTGTACTGGGTGCCCCAGCAGTGCCAGCCCGCCGGCGCTGTGCTCGCTCGATTTCTCAC
TGGGCCTTAGCAGCCTTCCCGCGGGGCAGGGCTCGGGACCTGCAGCCCGCCATGCCTGAGCCTCC
CCTCCATGGGCTCCTGTGCGGCCCGAGCCTCCCCGACGAGCACCACCCCCTGCTCCACAGCGCCC
AGTCCCATCGACCACGCAAGGGCTGAGAAGTGCGGGCGCACGGCACCGGGACTGGCAGGCAGCTA
CCCCTGCAGCCCTGGTGCGGAATCCACTGGGTGAAGCCAGCTGGGCTCCTGAGTCTGGTGGAGAC
TTGGAGAACCTTTATGTCTAGCTCAGGGATCGTAAATACACCAATCAGCACCCTGTGTCTAGCTC
AGGGTCTGTGAATGCACCAATCCACACTCTGTATCTAGCTACTCTGATGGGGCCTTGGAGAACCT
TTATGTCTAGCTCAGGGATTGTAAATACACCAATCGGCACTCTGTATCTAGCTCAAGGTTTGTAA

FIG. 5A

ACACACCAATCAGCACCCTGTGTCTAGCTCAGGGTATGTGAATGCACCAATCGACAGTCTGTATC
TGGCTACTTTCATGGGCATCCGTGTGAAGAGACCACCAAACAGGCTTTGTGTGAGCAATAAAGCT
TCTATCACCTGGGTGCAGGTGGGCTGAGTCCGAAAAGAGAGTCAGCGAAGGGAGATAAGGGTGGG
GCCGTTTTATAGGATTTGGGTAGGTAAAGGAAAATTACAGTCAAAGGGGGTTTGTTCTCTGGCGG
GCAGGAGTGGGGGGTCGCAAGGTGCTCAGTGGGGGTGCTTTTTGAGCCAGGATGAGCCAGGAAAA
GGACTTTCACAAGGTAATGTCATCAATTAAGGCAAGGACCCGCCATTTACACCTCTTTTGTGGTG
GAATGTCATCAGTTAAGTTGGGGCAGGGCATATTCACTTCTTTTGTGATTCTTCAGTTACTTCAG
GCCATCTGGGCGTATATGTGCAAGTTACAGGGGATGCGATGGCTTGGCTTGGGCTCAGAGGCTTG
ACAGCTACTCTGGTGGGGCCTTGGAGAATGTTTGTGTCGACACTCTGTATCTAGTTAATCTAGTG
GGGACGTGGAGAACCTTTGTGTCTAGCTCAGGGATTGTAAACGCACCAATCAGCGCCCTGTCAAA
ACAGACCACTCGGCTCTACCAATCAGCAGGATGTGGGTGGGGCCAGATAAGAGAATAAAAGCAGG
CTGCCCGAGCCAGCAGTGGCAACGCGCACAGGTCCCTATCCACAATATGGCAGCTTTGTTCTTTT
GCTGTTTGCGATAAATCTTGCTACTGCTCGCTTTTTGGGTCCACACTGCTTTTATGAGCTGTAAC
ACTCACCACGAAGGTCTGCAGCTTCACTCCTGAAGCCACTAAGACCACGAGCCCACCGGGAGGAA
TGAACAACTCCGGCCGCGCTGCCTTAAGAGCTATAACACTCACCGCGAAGGTCTGCAGCTTCACT
CCTCAGCCAGCGAGACCACGAACCCACCAGAAGGAAGAAACTGCGAACACATCTGAACATCAGAA
GGAACAAACTCCAGATGCACCACCTTAAGAGCTGTAACACTCACTGCGAGGGTCCGCGGCTTCCT
TCTTGAAGTCAGTGAGACCAAGCACTCACCAGTTTCGGACACAAGCCCAGGAGTTTGAGATCAGC
CTGGGCAACATGATGAAATGCCCTCTCTGCAAAAAAAAAAAAAAAATTACAAAAATTGGCGGAGCAT
GGTGGTCCGTGCCTGTGGTCCCAGCTACGCGGGAGGCTAAAGTGGGAGGATCGCTTGAGCCTGGG
AGGTGAAGACTGCAGTGAGCTGTGATTGTACCACAGCCCTCTAGGCTGGGGGACAGACTGAGACC
CTGTTTCCCCTCCGCAAAAAAATTGACAAAAGTGTAATAAGAGGTGCCTGATATGGCTAGGCGCA
GTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGCGGGCGGGTCACCTAAGGTCAGGA
GTGTGAGACCAGCCTGGCCAACATGGAGAAAGCCCATCTCTTCTAAAAATACAAAATTAGCCGGC
TGTGGGGGCAGTGGTGGAGCATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCA
CTTGAACCCAGGAGGCGGCGGTTGCAGTGAGCCGAGATCGTGCCATTGCACTCCACCCACTCCAG
CCTGGGCAACAAGAGCCAAACTCTGTCTTAAAAAAAAAAAAAAAAAGTGCCTGACATATAAGAGG
TGTGCAATGCAATAGTTGCCAGGCAACATGTTTAAGAATGTGGAGCTCCTGCCTTCCATGGTCCT
GTTAAAAACCCACCCTCAAGGCCAGGTGCAGTGGCTCATGCCTATAATCCCAGCACTTTGGGAGG
CCGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGACCACCAACATGGTGAAAT
CCCACCTCTACTAAAAATACAAAATTAGATGAGCATGGTGGTGCATGCCTGTAATCCCACCTACT
TGGGAGGCTGAGGCAGGAAAATCACTAGAACCAGGGAGGCGGAGGTTGTAGTGAGCCGAGATCGT
GCCATTGCACTCCAGCCTGAGCAATGAGCGAAACTCCATCTCAAAAAAACAACAACAAAAACCCA
CTCTCTACTCCCAGGGAGCTGGGTACAGAGCTGGGCCACATCAGTGCAAGGTGCTGAGCCACAGA
GCTAAGGCGGAGCTGCAGGACCGCGGACCAGATAACAGTGTGTGAGATCAGTGTGTGAGATCAGA
CGTCCCTGCCATTGGTGACCACCAGGGGGCCCCCAAGCACCAGAGATGGCCCCATCCAGTCACCA
CATCCACTTCTCATCCAGAGATGTCTGTTTCTTGGCACGCTGGGGTAAATTAGGACAGAAGGTGA
CAGTCTTGGGTGTGGTCAGTCAGACTGCCCCAGGCAGGCCTTGTGGCCTGTAGAAAACGTTCAGG
CCTAGGCCGGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGAT
CACGAGGTCAGGAGATCGTGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATAC
AAAAAATTGGCCGGGCATGGTGGCGGGCACCTGTAGTTCCAGCTACTCGGGAGGCTGAGGCAGGA
GAATGGCGTGAACCCGAGAGGCAGAGTTTGCAGTGAGCCGAGATCGCGCCACTGCACTCCAGCCT
GGGCGACAGAGCAAGACTCCATCTGGAAAAGAAAAAGAAAACGTTCAGGTCTGAGCCAGAGGCCC
AGGCTGTAATTCTGTCACTTACCATGACCTTGGGCAAGGCACTTCCTTCCCTGGCCCAGTTCACG
GGGTTGGAATCGACTCCAAGGTCCCTTCCAGCATTAACGCTGCATGGTTCTAAGATGAGAAGATG
GGGCAGTTTCCCCTCTCTCACCCCAGCCCGTGTCCACTTCAAGGTGAATGACCAGGGAAGTCACG
TGTCCCAATCCCGCAGTTCCAAAGCCCTTGGGGACCCTACTGTCAGGGTCGTGCACGAGGAGGTG
AAGGTCAGGTGAGCCAATCGCCTCGAAGGGTCTTGCCTCATTCGGGACAGACATCCGGTTTCCTC
TGGCTCTACCGGGATTCTAGGGGCTTTAGCCGAATGAGTCATGGGGGGCGGGGGGGTTTCTGGGG
GAGTTCCCAGCTAATCAACTTGGGACAGGACAGCCTGGAACTTTCGATGGTGCCTATCCAAGTGT
GGGGTGGGCACAGCAGCCAAGACCCAATGTCCTTATCTCAGGTAGGGGCTCAGGAGGTCTCCCAG
ACAGGCAGCCTCCGGAGAGTTTGGGGGTAGGAATGGGAGCAACCAGGCTTCTTTTTTTCTCTCTT
AGAATTTGGGGGCTTGGGGGACAGGCTTGAGAATCCCAAAGGAGAGGGGCAAAGGACACTCCCCC
ACAAGTCTGCCAGAGCGAGAGAGGGAGACCCCGACTCAGCTGCCACTTCCCCACAGGCCT

FIG. 5B

CC GGCAGTCCTC
ACAGCCCTCG TTCGCTCTCG GCGCCTCCTC TGCCTGGGCT CCCACTTCGG TGGCACTTGA
GGAGCCCTTC AGCCCACCGC TGCACTGTGG GAGCCCCTTT CTGGGCTGGC CAAGGCCAGA
GCCGGCTCCC TCAGCTTGCA GGGAGGTGTG GAGGGAGAGG CTCAAGCAGG AACCGGGGCT
GCGCACGGCG CTTGCGGGCC AGCTGGAGTT CCGGGTGGGC GTGGGCTTGG CGGGCCCCGC
ACTCGGAGCA GCGGGCCAGC CCTGCCAGGC CCCGGGCAAT GAGAGGCTTA GCACCCGGGC
CAGCGGCTGC GGAGGGTGTA CTGGGTGCCC CAGCAGTGCC AGCCCGCCGG CGCTGTGCTC
GCTCGATTTC TCACTGGGCC TTAGCAGCCT TCCCGCGGGG CAGGGCTCGG GACCTGCAGC
CCGCCATGCC TGAGCCTCCC CTCCATGGGC TCCTGTGCGG CCCGAGCCTC CCCGACGAGC
ACCACCCCCT GCTCCACAGC GCCCAGTCCC ATCGACCACG CAAGGGCTGA GAAGTGCGGG
CGCACGGCAC CGGGACTGGC AGGCAGCTAC CCCTGCAGCC CTGGTGCGGA ATCCACTGGG
TGAAGCCAGC TGGGCTCCTG AGTCTGGTGG AGACTTGGAG AACCTTTATG TCTAGCTCAG
GGATCGTAAA TACACCAATC AGCACCCTGT GTCTAGCTCA GGGTCTGTGA ATGCACCAAT
CCACACTCTG TATCTAGCTA CTCTGATGGG GCCTTGGAGA ACCTTTATGT CTAGCTCAGG
GATTGTAAAT ACACCAATCG GCACTCTGTA TCTAGCTCAA GGTTTGTAAA CACACCAATC
AGCACCCTGT GTCTAGCTCA GGGTATGTGA ATGCACCAAT CGACAGTCTG TATCTGGCTA
CTTTCATGGG CATCCGTGTG AAGAGACCAC CAAACAGGCT TTGTGTGAGC AATAAAGCTT
CTATCACCTG GGTGCAGGTG GGCTGAGTCC GAAAAGAGAG TCAGCGAAGG GAGATAAGGG
TGGGGCCGTT TTATAGGATT TGGGTAGGTA AAGGAAAATT ACAGTCAAAG GGGGTTTGTT
CTCTGGCGGG CAGGAGTGGG GGGTCGCAAG GTGCTCAGTG GGGGTGCTTT TTGAGCCAGG
ATGAGCCAGG AAAAGGACTT TCACAAGGTA ATGTCATCAA TTAAGGCAAG GACCCGCCAT
TTACACCTCT TTTGTGGTGG AATGTCATCA GTTAAGTTGG GGCAGGGCAT ATTCACTTCT
TTTGTGATTC TTCAGTTACT TCAGGCCATC TGGGCGTATA TGTGCAAGTT ACAGGGGATG
CGATGGCTTG GCTTGGGCTC AGAGGCTTGA CAGCTACTCT GGTGGGGCCT TGGAGAATGT

SalI
TTGTGTCGAC ACTCTGTATC TAGTTAATCT AGTGGGGACG TGGAGAACCT TTGTGTCTAG
CTCAGGGATT GTAAACGCAC CAATCAGCGC CCTGTCAAAA CAGACCACTC GGCTCTACCA
ATCAGCAGGA TGTGGGTGGG GCCAGATAAG AGAATAAAAG CAGGCTGCCC GAGCCAGCAG
TGGCAACGCG CACAGGTCCC TATCCACAAT ATGGCAGCTT TGTTCTTTTG CTGTTTGCGA
TAAATCTTGC TACTGCTCGC TTTTTGGGTC CACACTGCTT TTATGAGCTG TAACACTCAC
CACGAAGGTC TGCAGCTTCA CTCCTGAAGC CACTAAGACC ACGAGCCCAC CGGGAGGAAT
GAACAACTCC GGCCGCGCTG CCTTAAGAGC TATAACACTC ACCGCGAAGG TCTGCAGCTT

FIG. 6A

CACTCCTCAG CCAGCGAGAC CACGAACCCA CCAGAAGGAA GAAACTGCGA ACACATCTGA
ACATCAGAAG GAACAAACTC CAGATGCACC ACCTTAAGAG CTGTAACACT CACTGCGAGG
GTCCGCGGCT TCCTTCTTGA AGTCAGTGAG ACCAAGCACT CACCAGTTTC GGACACAAGC
CCAGGAGTTT GAGATCAGCC TGGGCAACAT GATGAAATGC CCTCTCTGCA AAAAAAAAAA
AAATTACAAA AATTGGCGGA GCATGGTGGT CCGTGCCTGT GGTCCCAGCT ACGCGGGAGG
CTAAAGTGGG AGGATCGCTT GAGCCTGGGA GGTGAAGACT GCAGTGAGCT GTGATTGTAC
CACAGCCCTC TAGGCTGGGG GACAGACTGA GACCCTGTTT CCCCTCCGCA AAAAAATTGA
CAAAAGTGTA ATAAGAGGTG CCTGATATGG CTAGGCGCAG TGGCTCATGC CTGTAATCCC
AGCACTTTGG GAAGCCGAGG CGGGCGGGTC ACCTAAGGTC AGGAGTGTGA GACCAGCCTG
GCCAACATGG AGAAAGCCCA TCTCTTCTAA AAATACAAAA TTAGCCGGCT GTGGGGGCAG
TGGTGGAGCA TGCCTGTAAT CCCAGCTACT CAGGAGGCTG AGGCAGGAGA ATCACTTGAA
CCCAGGAGGC GGCGGTTGCA GTGAGCCGAG ATCGTGCCAT TGCACTCCAC CCACTCCAGC
CTGGGCAACA AGAGCCAAAC TCTGTCTTAA AAAAAAAAAA AAAAAGTGCC TGACATATAA
GAGGTGTGCA ATGCAATAGT TGCCAGGCAA CATGTTTAAG AATGTGGAGC TCCTGCCTTC
CATGGTCCTG TTAAAAACCC ACCCTCAAGG CCAGGTGCAG TGGCTCATGC CTATAATCCC
AGCACTTTGG GAGGCCGAGG CGGGTGGATC ACCTGAGGTC AGGAGTTCGA GACCAGCCTG
ACCACCAACA TGGTGAAATC CCACCTCTAC TAAAAATACA AAATTAGATG AGCATGGTGG
TG

FIG. 6B

```
      CCTG TAATCCCACC TACTTGGGAG GCTGAGGCAG GAAAATCACT AGAACCAGGG
AGGCGGAGGT TGTAGTGAGC CGAGATCGTG CCATTGCACT CCAGCCTGAG CAATGAGCGA
AACTCCATCT CAAAAAAACA ACAACAAAAA CCCACTCTCT ACTCCCAGGG AGCTGGGTAC
AGAGCTGGGC CACATCAGTG CAAGGTGCTG AGCCACAGAG CTAAGGCGGA GCTGCAGGAC
CGCGGACCAG ATAACAGTGT GTGAGATCAG TGTGTGAGAT CAGACGTCCC TGCCATTGGT
GACCACCAGG GGGCCCCCAA GCACCAGAGA TGGCCCCATC CAGTCACCAC ATCCACTTCT
CATCCAGAGA TGTCTGTTTC TTGGCACGCT GGGGTAAATT AGGACAGAAG GTGACAGTCT
-1457 TGGGTGTGGT CAGTCAGACT GCCCCAGGCA GGCCTTGTGG CCTGTAGAAA ACGTTCAGGC
-1397 CTAGGCCGGG CACGGTGGCT CACGCCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGCGGGT
-1337 GGATCACGAG GTCAGGAGAT CGTGACCATC CTGGCTAACA CGGTGAAACC CCGTCTCTAC
-1277 TAAAAATACA AAAAATTGGC CGGGCATGGT GGCGGGCACC TGTAGTTCCA GCTACTCGGG
-1217 AGGCTGAGGC AGGAGAATGG CGTGAACCCG AGAGGCAGAG TTTGCAGTGA GCCGAGATCG
-1157 CGCCACTGCA CTCCAGCCTG GGCGACAGAG CAAGACTCCA TCTGGAAAAG AAAAAGAAAA
-1097 CGTTCAGGTC TGAGCCAGAG GCCCAGGCTG TAATTCTGTC ACTTACCATG ACCTTGGGCA
-1037 AGGCACTTCC TTCCCTGGCC CAGTTCACGG GGTTGGAATC GACTCCAAGG TCCCTTCCAG
 -977 CATTAACGCT GCATGGTTCT AAGATGAGAA GATGGGGCAG TTTCCCCTCT CTCACCCCAG
 -917 CCCGTGTCCA CTTCAAGGTG AATGACCAGG GAAGTCACGT GTCCCAATCC CGCAGTTCCA
 -857 AAGCCCTTGG GGACCCTACT GTCAGGGTCG TGCACGAGGA GGTGAAGGTC AGGTGAGCCA
 -797 ATCGCCTCGA AGGGTCTTGC CTCATTCGGG ACAGACATCC GGTTTCCTCT GGCTCTACCC
 -737 GGATTCTAGG GGCTTTAGCC GAATGAGTCA TGGGGGGCGG GGGGGTTTCT GGGGGAGTTC
                                                        XcmI
 -677 CCAGCTAATC AACTTGGGAC AGGACAGCCT GGAACTTTCG ATGGTGCCTA TCCAAGTG
```

FIG. 7